Historic, archived document

Do not assume content reflects current scientific knowledge, policies, or practices.

A56.9
R31
Cop. 3

LIBRARY
RD
NOV 7 - 1957
U. S. DEPARTMENT OF AGRICULTURE

Application of The Extreme Value Statistical Distribution

to annual precipitation and crop yields

FEBRUARY 1957 ARS 41-13
Agricultural Research Service

UNITED STATES DEPARTMENT OF AGRICULTURE

CONTENTS

Page

Introduction 1

Application of Theory 4

- General Relationships 4
- Plotting, computational, and fitting procedures . . . 4
- Application of plotting and fitting procedure 5
- Agreement between theory and observations 9

Discussion 9

- Utility of the extreme value theory 9
- An example of crop dependability as influenced by fallow or continuous cropping 11

Summary and Conclusions 13

Literature Cited 14

UNITED STATES DEPARTMENT OF AGRICULTURE
Agricultural Research Service
Soil and Water Conservation Research Branch

In Cooperation with

The Agricultural Experiment Stations of
Colorado, Montana, Oklahoma, Oregon, and Texas

APPLICATION OF THE EXTREME VALUE STATISTICAL DISTRIBUTION TO ANNUAL PRECIPITATION AND CROP YIELDS[1]

D. L. Brakensiek and A. W. Zingg[2]

INTRODUCTION

The success of a farm enterprise in arid regions often depends on the ability of the farmer to adapt dependable and profitable farming practices. Within such regions one of the primary climatic elements influencing crop production is precipitation. The objective of this study was to investigate a method to characterize annual precipitation and crop yields.

Evaluation of the dependability or conversely the risk associated with a crop at a given location in the past has often followed regression techniques. An example of this approach is found in Mathews and Brown (7).[3] In this instance, the procedure consisted of deriving a linear relationship between yields of crops and variables such as soil moisture or precipitation. An estimate of the frequency of occurrence of various yield levels for an area was then determined by applying the derived relationship to point rainfall data from that area. Characterization of annual precipitation amounts has proceeded little past the determination of means and coefficients of variations. However, Potter (10) has shown that the extreme value distribution may be used to characterize the occurrence of annual precipitation amounts.

One of the limitations of the regression-approach to determine the dependability of crop production lies in the rationale for the relationship between yields and soil moisture and/or rainfall. Yields are influenced by phenomena not necessarily related to soil moisture or precipitation conditions, e.g., animal or insect infestations, wind, or hail. Further, the utilization of this approach requires laborious tabulations and computations.

---

[1] Cooperative investigations by the Western Soil and Water Management Section, Soil and Water Conservation Research Branch, Agricultural Research Service, USDA; and the Colorado, Montana, Oklahoma, Oregon, and Texas Agricultural Experiment Stations.

[2] Agricultural Engineer and Soil Conservationist, respectively, Soil and Water Conservation Research Branch, Agricultural Research Service, USDA.

[3] Figures in parentheses refer to Literature Cited, page 14.

With the above considerations in mind, an improvement in technique for evaluating or characterizing the dependability of a crop or cropping practice or precipitation should possibly employ a procedure having the following properties:

i) Have a rational model.
ii) Give consideration of all factors which influence the phenomenon.
iii) Provide a method which is simple computationally.

During recent years the statistical distribution of the extreme values in a sample has been utilized with success in many diverse applications (5). As applied to the largest extreme, the fundamental theorem has been stated as follows (4):

> "In a set of N independent extremes $y_1, y_2, \ldots, y_N$, each being the extreme of one of N sets of n observations each of an unlimited, exponentially-distributed variable, as both N and n grow large the cumulative probability that any one of these N extremes will be less than any chosen quantity, y, approaches the double exponential expression
>
> $$F(y) = e^{-e^{-a(y-\hat{y})}}, \qquad [1]$$
>
> where $\hat{y}$ is the mode of the set."

From [1], the average return period of a quantity y being equalled or exceeded is $T_m = 1/(1-F(y))$. [2]

A crop yield in any one year at a particular location is an observable quantity and its magnitude is governed by the interaction of environmental components. Such components include soil properties, farming practices, precipitation, temperature, insect infestations and other factors. Conceptually then, the observed yield comes from a large population of possible values ranging from zero to some upper bound. It seems reasonable that the conceptual population of yields should be distributed approximately in an exponential manner, being bounded from below and above. Under dryland farming conditions most observed annual yields are far removed from the upper bound and, to a lesser extent, from the lower bound of zero. Recorded zero yields may be indicative of a crop not warranting harvest and in this event would not strictly be a zero yield. Thus, with reservations the conceptual population of annual yields may approximate the unlimited exponential distribution required by the fundamental theorem. The observed yield for a given location and year can be treated as a maximum.

The same reasoning applies to observed annual precipitation. The observed annual amount is from a large conceptual population and can be treated as a maximum. Here again the population may be bounded by zero and some unknown upper bound. However, the lower bound, for the observational period considered is rarely if ever attained. It is not known whether the upper bound is ever attained. Again it is hypothesized that the fundamental theorem is satisfied, possibly to a closer extent than with annual crop yields.

This study is an empirical investigation pertaining to the validity of applying extreme value theory to annual yields and precipitation amounts, i. e., whether the model is reasonable. Data from several locations were utilized for this purpose (1, 2, 6, 8, 9).

APPLICATION OF THEORY

GENERAL RELATIONSHIPS

Detailed information on fitting equation [1] to observed data can be found in (3, 4, 5). Equation [1] can be linearized by taking an iterated natural logarithm of both sides, i.e.,

$$z = -\ln(-\ln F(y)) = a(y - \hat{y}) \qquad [3]$$

$$y = \hat{y} + \frac{z}{a}. \qquad [4]$$

Introducing two theoretical quantities, see (4),

$$a = \frac{\sigma_N,}{S_y} \text{ and } \hat{y} = \bar{y} - S_y \frac{\bar{y}_N}{\sigma_N},$$

equation [4] becomes

$$y = \bar{y} + S_y \frac{(z - \bar{y}_N)}{\sigma_N}. \qquad [5]$$

Dividing [5] by $\bar{y}$, one has

$$\frac{(y)}{\bar{y}} = 1 + \left(\frac{S_y}{\bar{y}}\right) \frac{(z - \bar{y}_N)}{\sigma_N}, \qquad [6]$$

where

$$\frac{(S_y)}{\bar{y}} = \text{coefficient of variation of the series of extremes, C. V.}$$

z is defined by [3], σ_N and $\bar{y}_N$ are tabulated quantities (4), see Table 2.

PLOTTING, COMPUTATIONAL, AND FITTING PROCEDURES

For an examination of how the theory represents the observed series, it is convenient to plot the data and to fit equation [6] to the plotted points. The method of procedure is as follows:

(a) The set of $\frac{(y)}{\bar{y}}$ values are ordered according to magnitude with the largest value first, the next largest value second,-----, the smallest value in the N^{th} position.

(b) Assign a rank, m, to the above ordered series, i. e., the first value has a rank of one, the next largest value a rank of two, -----, and the smallest value a rank of N.

(c) For each ordered value calculate $F\left(\frac{y}{\bar{y}}\right)$ as follows:

$$F\left(\frac{y}{\bar{y}}\right) = \frac{N + 1 - m}{N + 1}, \qquad [7]$$

where m is the rank of a particular $\left(\frac{y}{\bar{y}}\right)$ value and N is the total of observations.

If special probability paper is available (5) the points $(y/\bar{y},\ F\,(y/\bar{y}))$ are plotted. However, if such paper is not available it can be constructed or values of $(y/\bar{y},\ -\ln(-\ln F\,(y/\bar{y})))$ may be plotted on linear coordinate paper.

Calculation of the quantity $(S_y/\bar{y})$ is carried out with either the ranked or unranked data. Thus,

$$(S_y/\bar{y})^2 = \frac{1}{N-1} \quad \sum (y/\bar{y})^2 \quad - \quad \frac{N}{N-1}$$

$$\text{and} \quad C.\ V. = S_y/\bar{y} = \sqrt{\frac{1}{N-1} \quad \left(\sum (y/\bar{y})^2 - N \right)}.$$

For fitting equation [6] to the data it is convenient to utilize determined values of $\bar{y}_N$ and σN, the reduced means and standard deviations, respectively, for obtaining a "least squares" type fit. Values associated with N = 20 to N = 59, inclusive, are given in Table 2. Two points of equation [6], utilizing Table 2, can be located on the graph of $(y/\bar{y})$, $F\,(y/\bar{y})$ for orienting the straight line.

APPLICATION OF PLOTTING AND FITTING PROCEDURE

A detailed example of the plotting and fitting procedures will be of value for possible applications. For this purpose 47, (N = 47), years of annual precipitation data (Jan. 1 - Dec. 31) from Akron, Colo., were used, see Table 1. Values shown in Table 1 for each of the ranked precipitation amounts were obtained as follows:

$$N = 47$$

$$\bar{P} = \sum P / N = \frac{803.6}{47} = 17.097 \text{ ins.}$$

$$F\left(\frac{P}{\bar{P}}\right) = \frac{N + 1 - m}{N + 1} = \frac{48 - m}{48}, \quad (m = 1, \ldots\ldots, 47)\ .[7]$$

Calculation of $S_{P/\bar{P}}$, coefficient of variation, C. V., is found as follows:

$$\sum (P/\bar{P})^2 = 49.357861.$$

$$(S_{P/\bar{P}})^2 = \frac{49.357861}{46} - \frac{47}{46}$$

$$(S_{P/\bar{P}}) = \text{C. V.} = 0.2264$$

Thus the fitted equation is

$$\left(\frac{P}{\bar{P}}\right) = 1 + 0.2264 \left(\frac{z - \bar{y}_N}{\sigma_N}\right) \qquad [6]$$

With N = 47, values of $\bar{y}_N$ and σ_N from Table 2 are found to be 0.5473 and 1.1557, respectively. Two points on equation [6], chosen arbitrarily, can now be found as follows:

Assume that $F\,(^P/_{\bar{P}})_1 = 0.990,$

then $z_1, = -\ln(-\ln(0.990)\ ) = 4.60517$

and assume that $F\,(^P/_P)_2 = 0.800,$

then $z_2 = -\ln\ (-\ln(0.800) = 1.50058.$

Substituting the values of $\bar{y}_N$, and σ_N, and the calculated z_1 and z_2 values into equation [6] gives the following:

$$\left(\frac{P}{\bar{P}}\right)_1 = 1 + 0.2264 \left(\frac{4.60517 - 0.5473}{1.1557}\right)$$

$$\left(\frac{P}{\bar{P}}\right)_1 = 1.7949$$

and

$$\left(\frac{P}{\bar{P}}\right)_2 = 1. + 0.2264 \left(\frac{1.50058 - 0.5473}{1.1557}\right)$$

$$\left(\frac{P}{\bar{P}}\right)_2 = 1.1873$$

The "least squares" equation will pass through the points

$$\left(\left(P/\bar{P}\right)_1, \ F\left(\frac{P}{\bar{P}}\right)_1\right) \quad \text{and} \quad \left(\left(P/\bar{P}\right)_2, \ F\left(\frac{P}{\bar{P}}\right)_2\right),$$

i. e., (1.7949, 0.990) and (1.1873, 0.800).

TABLE 1.--ANNUAL PRECIPITATION DATA, AKRON, COLO.

P ins.	$\frac{P}{\bar{P}}$	Rank m	$F(\frac{P}{\bar{P}})$
26.5	1.550	1	0.979
25.0	1.462	2	0.958
23.1	1.351	3	0.937
22.5	1.316	4	0.916
22.3	1.304	5	0.895
22.1	1.293	6	0.875
21.6	1.263	7	0.854
21.4	1.251	8	0.833
20.7	1.211	9	0.812
20.4	1.193	10	0.791
20.2	1.181	11	0.771
19.6	1.146	12	0.750
19.1	1.117	13	0.729
19.0	1.111	14	0.708
18.9	1.105	15	0.687
18.7	1.094	16	0.666
18.5	1.082	17	0.646
18.2	1.064	18	0.625
18.2	1.064	19	0.604
18.1	1.058	20	0.583
17.7	1.035	21	0.562
17.5	1.023	22	0.542
17.4	1.017	23	0.521
16.8	0.982	24	0.500
16.7	0.976	25	0.479
16.7	0.976	26	0.458
16.6	0.971	27	0.437
15.7	0.918	28	0.416
15.6	0.912	29	0.395
15.6	0.912	30	0.374
15.5	0.906	31	0.353
15.5	0.906	32	0.332
15.4	0.900	33	0.311
15.1	0.883	34	0.290
14.5	0.848	35	0.269
14.4	0.842	36	0.248
14.1	0.825	37	0.227
13.7	0.801	38	0.206
13.4	0.784	39	0.185
12.7	0.743	40	0.164
12.6	0.737	41	0.143
11.8	0.690	42	0.122
11.7	0.684	43	0.101
11.4	0.666	44	0.080
11.0	0.643	45	0.059
10.7	0.626	46	0.038
9.7	0.567	47	0.017

TABLE 2.--REDUCED MEANS AND STANDARD DEVIATION[1]

N	$\bar{y}_N$	σ_N	N	$\bar{y}_N$	σ_N
20	0.5236	1.0628	40	0.5436	1.1413
21	0.5252	1.0696	41	0.5442	1.1436
22	0.5268	1.0754	42	0.5448	1.1458
23	0.5283	1.0811	43	0.5453	1.1480
24	0.5296	1.0864	44	0.5458	1.1499
25	0.5309	1.0915	45	0.5463	1.1519
26	0.5320	1.0961	46	0.5458	1.1538
27	0.5332	1.1004	47	0.5473	1.1557
28	0.5343	1.1047	48	0.5477	1.1574
29	0.5353	1.1086	49	0.5481	1.1590
30	0.5362	1.1124	50	0.5485	1.1607
31	0.5371	1.1159	51	0.5489	1.1623
32	0.5380	1.1193	52	0.5493	1.1638
33	0.5388	1.1226	53	0.5497	1.1658
34	0.5396	1.1255	54	0.5501	1.1667
35	0.5403	1.1285	55	0.5504	1.1681
36	0.5410	1.1313	56	0.5508	1.1696
37	0.5418	1.1339	57	0.5511	1.1708
38	0.5424	1.1363	58	0.5515	1.1721
39	0.5430	1.1388	59	0.5518	1.1734

---

[1] Table from p. 65 (4) Used with the permission of the authors and publishers.

AGREEMENT BETWEEN THEORY AND OBSERVATIONS

Figure 1 shows the accumulated probability $F\ (^{P}/_{\bar{P}})$ and $F\ (^{Y}/_{\bar{Y}})$ corresponding to levels of $(^{P}/_{\bar{P}})$ and $(^{Y}/_{\bar{Y}})$ at selected locations. Shown in this figure are the plottings of both precipitation and crop data from 6 locations for given lengths of record, N. Average precipitation, $\bar{P}$, is in inches and yields are in bushels per acre except cotton which is in pounds per acre. Coefficients of variation, C. V., are also given for each series of data. The broken lines on each plot indicate the "50 per cent" confidence bands. Procedures for construction of these bands can be found in (3, 4, 5).

It has been stated (4) that if two-thirds of the observations, as plotted on extreme value probability paper, fall within the confidence band, the observed series may be considered to be represented adequately by the theory of extreme values. For the situations studied here this criterion is satisfied. As would be deduced from the fundamental theorem, the series of continuous wheat yields in which zero values are encountered indicated the poorest fit. Conversely, for situations in which zero values seldom, if ever, occur, i.e., fallow wheat yields and annual precipitation, a better fit was evident.

DISCUSSION

UTILITY OF THE EXTREME VALUE THEORY

The simplicity of the present technique is evident in that only the quantity $\frac{Sy}{\bar{y}}$ and plotting positions need to be calculated from the data. Construction of special probability paper makes the plotting procedure simple.

Having a description of the past population of crop yields or precipitation amounts, it would be possible, if the observed population is characteristic of the larger parent population, to predict selected crop or precipitation characteristics with an associated probability statement. It should be pointed out that the "classical" approach such as deriving a relation between yields and precipitation in order to predict future yields must utilize some prediction or probability description of future precipitation. With the present approach the intermediate regression step does not exist as an estimate can be made directly from the distribution of past yields wherein all factors influencing yield have manifested themselves.

An examination of January to December precipitation and July to June precipitation led to the hypothesis that any annual total would yield similar results. Further, monthly and seasonal amounts of precipitation were found amenable to description by the technique.

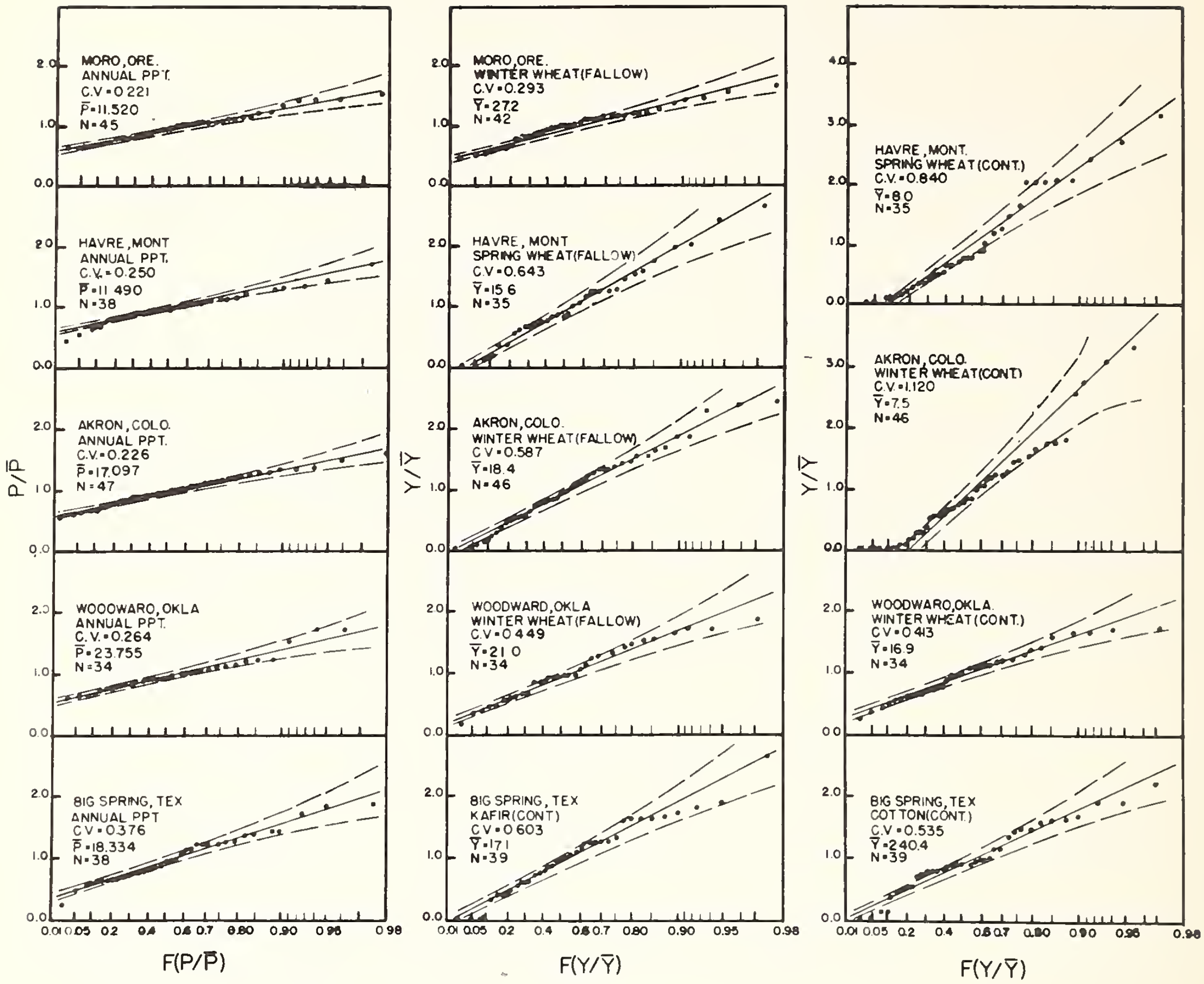

Figure 1. Relationship between accumulative probability, F, and precipitation and yield ratios of annual values to long time means.

AN EXAMPLE OF CROP DEPENDABILITY AS INFLUENCED BY FALLOW OR CONTINUOUS CROPPING:

The dependability of a crop can be measured by the frequency of occurrence of yields, for a selected length of record, within arbitrary intervals, e.g., $Y < 5$ bu./ac., 5 bu./ac. $\leq Y < 15$ bu./ac. From the development presented such determinations are readily obtained by the following expression,

$$\Pr \{ L \leq Y < U \} \equiv \Pr \{ Y < U \} - \Pr \{ Y < L \}$$

where Pr is the probability of the yield level Y falling between arbitrary levels L and U. Each of the expressions on the right-hand side are directly read from the pertinent graph of Figure 1.

As an example of evaluating crop dependability, winter wheat yields at Akron, Colo., and Woodward, Okla., are contrasted with respect to a continuous and summer fallow cropping practice.

In order to evaluate yield interval probability statements utilizing Figure 1 the interval limits will have to be changed to ones for the ratio $Y/\bar{Y}$. For example at Akron, the average $\bar{Y}$, for continuous wheat is 7.532 bu./ac. Thus values of Y equal to 5 bu./ac. and 15 bu./ac. correspond to values of $Y/\bar{Y}$ of 0.664 and 1.991, respectively. Probability values are derived as follows:

$$\Pr \{ Y < 5 \} \equiv \Pr \{ Y/\bar{Y} < 0.664 \} = 0.44$$

$$\Pr \{ 5 < Y \leq 15 \} \equiv \Pr \{ 0.664 < Y/\bar{Y} \leq 1.991 \} =$$

$$\Pr \{ Y/\bar{Y} < 1.991 \} - \Pr \{ Y/\bar{Y} < 0.664 \} = 0.81 - 0.44 = 0.37$$

An examination of Table 3 clearly shows why fallow cropping is preferred at Akron and continuous cropping is preferred at Woodward. It must be remembered that fallow cropping cuts the acreage harvested annually in half. From the first line of Table 3 it will be noted that the probability of obtaining a yield less than 5 bu./ac. at Akron for continuous cropping is 0.44. With fallow wheat it is 0.09. The hazard of crop failure, therefore, is almost 5 times greater with continuous wheat. By contrast, at Woodward, the probability of obtaining a crop failure is at the low level of 0.02 for both methods of cropping.

TABLE 3.--PROPORTION OF THE TIME WHEAT YIELDS FALL IN VARIOUS YIELD INTERVALS

YIELD INTERVAL	AKRON, COLO.		WOODWARD, OKLA.	
	Continuous	Fallow	Continuous	Fallow
Bu. per acre				
< 5	0.44	0.09	0.02	0.02
≥ 5 < 15	0.37	0.35	0.41	0.30
≥ 15 < 25	0.14	0.31	0.42	0.38
≥ 25 < 35	0.03	0.16	0.12	0.19
≥ 35	0.02	0.09	0.03	0.11

SUMMARY AND CONCLUSIONS

For the data considered in this study the extreme value theory satisfactorily represents the distribution of annual crop yields and annual precipitation amounts. The computational and graphical procedure is simple. It is possible to characterize selected crop or precipitation measures by their probability of occurrence.

The extreme value technique presented herein appears to offer a technique adapted for an extensive survey of cropping risks, as well as expected precipitation amounts. A study of this type would be useful in depicting the hazards of crop production in areas such as the Great Plains. For some purposes there may be advantages in working directly with the distribution of crop yields and allowing all factors which influence them to be taken into account.

The marked difference between the annual precipitation and yield graphs indicates that annual precipitation for the situations studied here is not a satisfactory indicator of the variability of crop yields. Based on the levels of the coefficient of variation, annual precipitation showed less variability than annual yields.

LITERATURE CITED

(1) ARMY, T. J. (Unpublished precipitation and crop data for Havre, Mont.)

(2) BRANDON, J. F. and B. W. GREB. (Unpublished precipitation and crop data.) U. S. Akron Field Station, Akron, Colo.

(3) CHOW, V. T. Frequency analysis of hydrologic data with special application to rainfall intensities. Univ. of Illinois Engr. Exp. Sta. Bul. Series 414, 1953.

(4) COURT, ARNOLD. Some new statistical techniques in geophysics. In Landsberg, H. E., and others, Advances in Geophysics, Vol. 1, pp. 45-85. Academic Press, Inc., New York, 1952.

(5) GUMBEL, E. J. Statistical theory of extreme value and some practical applications. Nat. Bur. of Stds. App. Math. Series 33, 1954.

(6) LOCKE, L. F. and O. R. MATHEWS. Relations of cultural practices to winter wheat production, Southern Great Plains Field Station, Woodward, Okla. U. S. Dept. Agr. Cir. 917, 1953.

(7) MATHEWS, O. R. and L. A. BROWN. Winter wheat and sorghum production in the Southern Great Plains under limited rainfall. U. S. Dept. Agr. Cir. 477. 1938.

(8) MOLDENHAUER, W. S., and F. E. KEATING. A study of relationships between climatic factors and yields of cotton, milo, and kafir at Big Spring, Texas. U.S.D.A., Soil and Water Conservation Res. Branch Res. Rpt. No. 295. 1956.

(9) OVESON, M. M., and W. E. HALL. Long time wheat-land tillage experiment on dryland soils of Eastern Oregon. (Unpublished.)

(10) POTTER, W. D. Normalcy tests of precipitation and frequency studies of runoff on small watersheds. U. S. Dept. Agr. Tech. Bull. 985. 1949.